RESEARCH ARTICLE OPEN ACCESS

Functional Principal Component Analysis as an Alternative to Mixed-Effect Models for Describing Sparse Repeated Measures in Presence of Missing Data

Corentin Ségalas[1] | Catherine Helmer[2] | Robin Genuer[1] | Cécile Proust-Lima[2]

[1]Univ. Bordeaux, INSERM, INRIA, BPH, U1219, Bordeaux, France | [2]Univ. Bordeaux, INSERM, BPH, U1219, Bordeaux, France

Correspondence: Corentin Ségalas (corentin.segalas@u-bordeaux.fr)

ABSTRACT

Analyzing longitudinal data in health studies is challenging due to sparse and error-prone measurements, strong within-individual correlation, missing data and various trajectory shapes. While mixed-effect models (MM) effectively address these challenges, they remain parametric models and may incur computational costs. In contrast, functional principal component analysis (FPCA) is a non-parametric approach developed for regular and dense functional data that flexibly describes temporal trajectories at a potentially lower computational cost. This article presents an empirical simulation study evaluating the behavior of FPCA with sparse and error-prone repeated measures and its robustness under different missing data schemes in comparison with MM. The results show that FPCA is well-suited in the presence of missing at random data caused by dropout, except in scenarios involving most frequent and systematic dropout. Like MM, FPCA fails under missing not at random mechanism. The FPCA was applied to describe the trajectories of four cognitive functions before clinical dementia and contrast them with those of matched controls in a case-control study nested in a population-based aging cohort. The average cognitive declines of future dementia cases showed a sudden divergence from those of their matched controls with a sharp acceleration 5 to 2.5 years prior to diagnosis.

1 | Introduction

In medical research, it is very common to collect in a cohort of participants the repeated measures of various markers at different visits in order to study the trajectories of the underlying biological processes. For example, in HIV research studies, the temporal evolution of CD4 is used to assess the immune system response [1]. In prostate cancer, the prostate specific antigen is a marker used to monitor prostate cancer progression [2]. In cognitive ageing studies, the temporal evolution of neuropsychological tests is used to assess the cognitive abilities trend among the older adults [3].

Marker data repeatedly measured over time in longitudinal studies have several characteristics that raise analytical challenges. First, the structure of the data induces a strong within-individual correlation which implies a variability across individuals. Second, the data collection process [4], often heavy in medical studies, usually leads to sparse measurements and individual-specific timings. For instance, in epidemiological cohorts of chronic diseases, it is not rare to have data every 2 to 5 years only [1–3]. Third, participants may leave the study before the end because they died or because they refused to undergo the next visit which leads to dropout [5]. These early truncations of the longitudinal process are often linked to the health status of the

---

participant, and thus are likely informative. Fourth, marker data collected in health studies are systematically noisy observations of the true underlying process of actual interest. Finally, the marker trajectories may exhibit different shapes, requiring flexible modeling tools of time trends. Figure 1 illustrates the characteristics of such marker data in the context of four neuropsychological tests collected in a cognitive aging study with the error-prone trajectories and sparse measurements (left panels), the individual-specific timings (middle panels), and the missingness patterns (right panels).

Mixed-effect models (MM) are the standard method to analyze longitudinal data in health studies [6]. They extend the classic linear regression models to correlated data by modeling both marginal trajectories using fixed effects and the individual-specific deviations using individual regression coefficients called the random effects. MM split the observations into the observation error and the underlying data generation process defined in continuous time, thus handling any measurement time and accounting for serial correlation. Any basis of time functions can be considered, including splines or polynomials, for flexible modeling. As estimated in the maximum likelihood framework, MM estimates are robust to missing data as long as the missingness mechanism is predictable by the observed

data (mechanism also known as missing at random) [7, 8]. MM thus handle most challenges faced with the longitudinal data collected in health data. However, MM are parametric tools; the distribution of both the error and the random effects must be specified by the user. Also, in some specific cases, the flexibility put into the MM may come with a great computational cost, particularly when the model is nonlinear and includes many random effects because of an heavy numerical integration in the log-likelihood calculation [9].

Repeated measures of markers are observations of a continuous-time process and as such they could also be considered as sparse and irregular functional data. Indeed, classic functional data is a collection of noisy observations from an unknown random function f on a regular and temporally dense grid of time. In functional data analysis, one of the main goal is to describe and summarize the temporal evolution of these trajectories. One of the most well-known method to do so is the functional principal component analysis (FPCA) [10, 11]. Similarly to classic principal component analysis, it projects functional data into a lower dimensional functional space defined by a functional basis. In this functional space, each participant has its own set of coordinates called functional principal component scores. In health studies, FPCA is also very appealing because it allows,
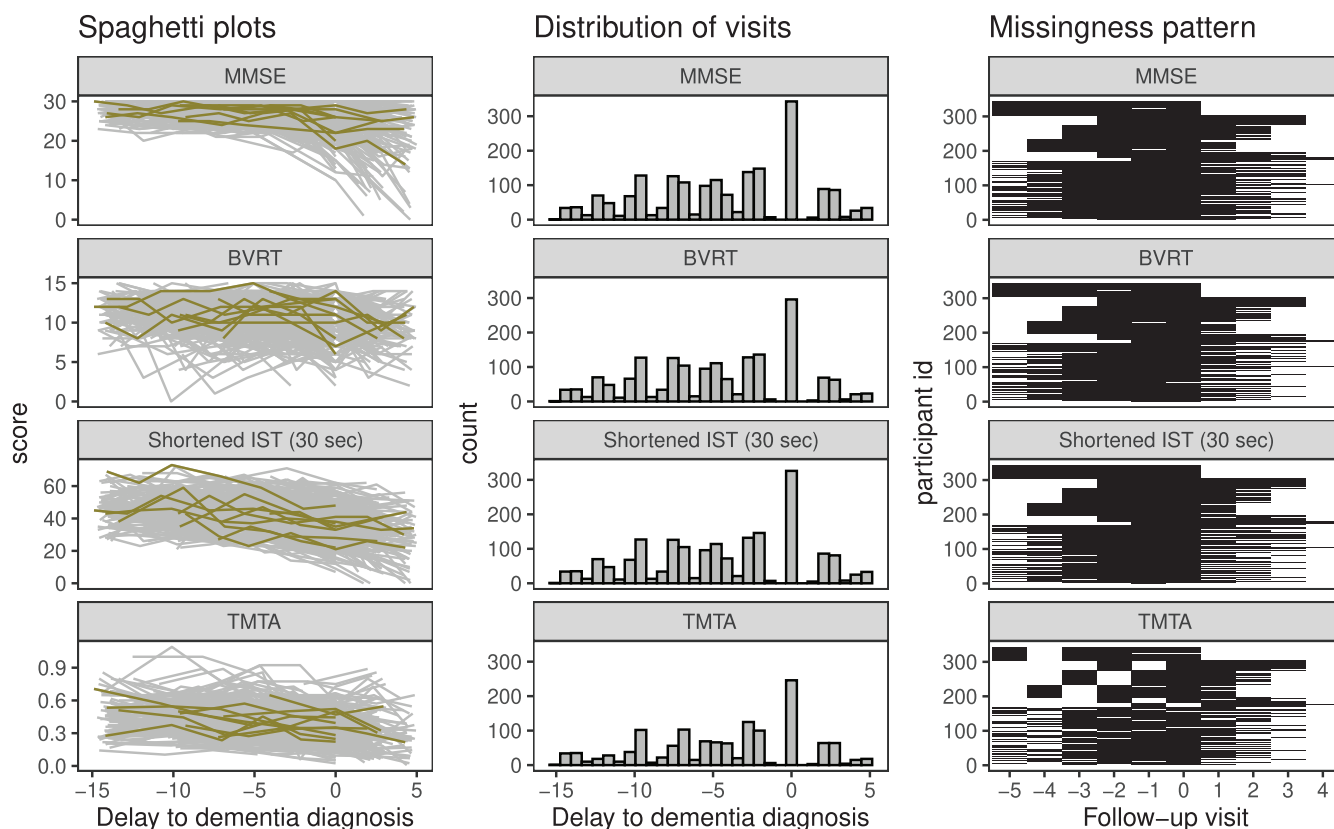

FIGURE 1 | Description of the repeated data in 174 dementia cases and 174 matched controls ($N = 348$) from the 3C study according to the delay (in years) to dementia diagnosis for the cases, and the delay to the dementia diagnosis of the matched case for controls (with 0 corresponding to the dementia diagnosis visit of the case). Are reported the individual observed trajectories over time to diagnosis of four cognitive scores (mini-mental state examination [MMSE], Benton visual retention test [BVRT], Isaacs set test [IST30], and trail making test score part A [TMTA]) with 10 randomly selected individual trajectories emphasized in green (first column); the distribution of the times of measurement (second column) and the missingness pattern visualization for each individual at each visit (third column). A black segment means the score is observed, a white segment means no observation is available. The strong structure on the missingness pattern is a consequence of the matched nature of the data.

with a fully non parametric approach, to describe and summarize temporal trajectories in a parsimonious way using only a few principal functional components and individual contributions to these components. However, to be suitable for the analysis of longitudinal data stemmed from health studies, functional analysis methods should necessarily handle the challenges of sparsity, irregularity of the observation grid and missing data. Previous work have proposed FPCA for sparse longitudinal data [12]. Furthermore, FPCA should work in theory with irregular observation grids although current implementations in R packages are not always flexible enough. The main unsolved challenge is the behavior of FPCA techniques with missingness mechanisms. No theoretical or numerical evaluation has been sought. While FCPA is expected to behave well in case of completely random missingness since it constitutes a specific type of irregular grid, the behavior of the FPCA to other types of missingness still deserve to be investigated, especially when working with longitudinal data from health studies where informative missingness is common. The aim of this article is to evaluate the robustness of FPCA to various schemes of missing data in an empirical simulation study and to compare its performances to the classic mixed-effect model approach.

Section 2 briefly introduces the motivating data example with repeated evaluations of cognitive functioning in a population-based aging cohort. In Section 3, the statistical framework–in particular the missing data mechanisms–is described and both the MM and the FPCA are introduced. Results from two empirical simulation studies are reported in Section 4. The two methods are then applied in Section 5 to describe the trajectories of four cognitive functions in years preceding clinical dementia onset in contrast with the trajectories of matched controls. Our findings are summarized and discussed in Section 6.

2 | The Three-City Study

The three-city study (3C study) is a French observational cohort study started in 1999 aiming at exploring the link between vascular diseases and dementia in the older adults [3]. Subjects aged 65 years and older at baseline were recruited in three French cities and followed-up every 2 to 3 years during 17 years. At each visit, the cognitive functioning was measured by a battery of cognitive tests, and a diagnosis of dementia was established according to a three steps procedure: a screening based on neuropsychological performance; a clinical evaluation performed by a neurologist; and a final diagnosis made by an independent committee of neurologists (see [3] for details).

Our motivating aim was to describe the cognitive trajectories of participants diagnosed with dementia and contrast them with those expected during natural aging. To this end, we built a nested case-control study from the 3C study. We included the 174 incident dementia cases diagnosed in the Bordeaux Center with at least one measure at each cognitive test, and individually matched them in density of incidence with 156 controls free of dementia at the time of diagnosis of the case. The matching was done according to sex, presence of at least one APOE-ϵ4 allele (main genetic risk factor of dementia), level of education, age at diagnosis with a 3-year margin, and length of follow-up at diagnosis (in years). We removed the cognitive measures occurring

5 years after the diagnosis (considered as time 0) as most cases were not able anymore to undergo a large cognitive battery.

We focused on four psychometric scores that evaluate different cognitive aspects: the mini-mental state examination score (MMSE) [13] which evaluates global cognitive functioning (with range 0–30), the Benton visual retention test score (BVRT) [14] which measures visual memory (with range 0–15), the 30-s Isaacs set test (IST) score [15] which assesses verbal fluency (with range 0–73) and the number of correct moves per minute at the trail making test score part A (TMTA) [16] which assesses executive functioning and attention. Note that the TMTA was not assessed at the second visit.

Figure 1 describes the sample with the individual cognitive trajectories, the timings and patterns of missingness, thus illustrating the sparsity, irregularity in timings and missing data usually encountered in cohort studies.

3 | Methods

3.1 | Sparse and Irregular Repeated Measures Data

We assume that we observe repeated measures of a longitudinal marker over time t for each unit i of a sample of size N. These observations y_{ij} also noted $y_i(t_{ij})$ with $j = 1, \ldots, n_i$ lead to an individual vector of observations $\mathbf{y_i} = (y_{i1}, \ldots, y_{in_i})^\top$ where n_i denotes the number of observations for unit i. The observations y_{ij} are measurements with error of the underlying process of interest $y_i^\star(t)$ so that $y_{ij} = y^\star(t_{ij}) + \varepsilon_{ij}$ where ε_{ij} denotes the independent random observation error.

Classically, in the functional data framework, the observation grid $\mathbf{t} = (t_1, \ldots, t_n)^\top$ is dense, regular over time (e.g., hourly measures, daily measures) and the same across the n units. In contrast, in health studies on humans (cohorts, clinical trials), the data collection is frequently done with a lower frequency and the visit times $\mathbf{t_i} = (t_{i1}, \ldots, t_{in_i})^\top$ may vary across individuals. For example, in the 3C cohort data, there may be variations of several months or even years across individuals around the theoretical visits every 2 to 3 years because collecting cognitive test results implies an heavy and expensive observation process. Additionally, it is very common to have missing data which leads to different size of observation grid. Such a design defines sparse and irregular functional data.

3.2 | Missing Data Mechanisms

We use the classical formalism of missing data for a marker of interest as introduced by Little and Rubin [8]. We denote Y the set of all marker data which includes those that are observed Y^o and those that are missing Y^m. The binary variable R is the indicator variable of observation with $R_{ij} = 1$ if Y_{ij} is observed, and 0 otherwise. Little and Rubin [8] distinguishes three mechanisms of missingness. Data is called missing completely at random (MCAR) when the probability that the marker is missing does not depend on the marker at all,

$$P(R|Y^o, Y^m) = P(R)$$

It happens for instance when a cognitive test is unavailable at a certain visit because this test was not scheduled in the protocol for this visit (as this is the case for the TMTA at the second visit in 3C). Data is called missing at random (MAR) when the probability that the marker is missing depends on the observed values of the marker,

$$P(R|Y^o, Y^m) = P(R|Y^o)$$

It happens for instance when a participant refusal to pass a cognitive test depends upon his/her results at previous visits. Finally, data is called missing not at random (MNAR) when it can also depend on the unobserved values of the marker,

$$P(R|Y^o, Y^m) = P(R|Y^o, Y^m)$$

This can happen for instance when a participant refuses to pass the test because he/she feels that his/her cognitive level has dropped since last visit.

Missing data can be divided into two main types: intermittent missing data, when a measurement is missing during the follow-up but other measurement can be collected later on, and dropout data which happens when any visit after the dropout visit will be missing. In this work, we are interested in both types of missing data.

3.3 | Mixed-Effect Models

Mixed-effect models have been introduced by Laird and Ware [17] and since then have been widely used to model repeated measures collected in longitudinal health studies. We focus on a mixed model that assumes a continuous marker with zero-mean independent Gaussian errors ε_{ij} so that $y_{ij} = y_i^\star(t_{ij}) + \varepsilon_{ij}$. The underlying value of the marker at time t is then modeled as follows:

$$y_i^\star(t) = g(t, X_i(t), \boldsymbol{\beta}, \boldsymbol{b_i}) \tag{1}$$

where $X_i(t)$ denotes a set of time-dependent covariates, $\boldsymbol{\beta}$ denotes the vector of fixed parameters, $\boldsymbol{b_i}$ denotes the vector of random effects with $\boldsymbol{b_i} \sim \mathcal{N}(0, \boldsymbol{B})$ and g a basis of functions that describes the shape of the marker trajectory. The observation error $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ is assumed to be independent from the vector of random effects. In the remainder of this work, we concentrate on linear mixed-effect models where $g(t, X_i(t), \boldsymbol{\beta}, \boldsymbol{b_i}) = \sum_{k=1}^p (\beta_k + b_{ik}) g_k(t, X_i(t))$ with $(g_1, \dots, g_p)$ a p-basis of a priori specified functions of time, most often splines or polynomials.

Mixed-effect models have been extended to handle different types of data: binary or count data using generalized linear mixed model theory, or continuous data with strong ceiling or floor effect which is very common among neuropsychological scores [18, 19]. For the latter, a parametric monotonic function is applied to Y in Equation (1) to normalize the marker data.

The mixed-effect model is generally estimated by maximizing the log-likelihood:

$$\ell(\theta) = \sum_{i=1}^N \log \int f(Y_i|\boldsymbol{b_i}) f(\boldsymbol{b_i}) d\boldsymbol{b_i}$$

where θ denotes the model parameters ($\theta = (\boldsymbol{\beta}, \text{vec}(\boldsymbol{B}), \sigma)^\top$), $f(Y_i|\boldsymbol{b_i})$ is the density function of the outcome conditional to the random effects and $f(\boldsymbol{b_i})$ is the density of the random effects. When the model is linear in the random effects, the integral has an analytical solution. In other cases, the log-likelihood computation often implies to numerically approximate the integral [20].

Fitted values of existing observations or predictions of new data are easily derived from a fitted mixed model using the a posteriori distribution of the random effects given the individual observations. In linear mixed-effect models, it is directly computed using conditional expectation properties of multivariate normal variables while for nonlinear models, it is approximated by the mode of the posterior distribution [17, 21].

Despite the possible use of flexible families of time-functions (e.g., splines), a MM remains a parametric model with a distribution for the random effects and for the random error to be chosen (often Gaussian). Additionally, some assumptions on the variance structure of the random effect might be needed when the number of random effects increases to avoid a too large number of covariance parameters.

One essential feature of the MM is its robustness to MAR data, which ensures an unbiased estimator under MAR [7]. When the missing data are suspected to be MNAR, that is when the missing mechanism also depends on unobserved data, a joint model of the missingness mechanism is required. This can be done with a logistic model for intermittent missing data or a survival model for dropout [22, 23]. However, in any case, the model requires a correct specification of the missing data mechanism to yield robust estimates [24]. For instance, if the instantaneous risk of dropout is assumed to be associated with the marker level, the following proportional hazard model can be jointly estimated with the mixed model in (1):

$$\alpha_i(t) = \alpha_0(t) \exp(\bar{X}_{ij}^\top \boldsymbol{\gamma} + y_i^\star(t)^\top \boldsymbol{\eta})$$

with $\alpha_0(t)$ the baseline risk function, $\bar{X}_{ij}^\top$ a set of covariates, $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ two vectors of parameters to be estimated.

3.4 | Functional Principal Component Analysis

Functional principal component analysis has been introduced by Besse and Ramsay [25] and since then has been used for describing, modeling, predicting or classifying abundant temporal series [11, 26]. As for mixed-effect models, we consider additive random errors such that observations $y_{ij} = y_i^\star(t_{ij}) + \varepsilon_{ij}$ with $y_i^\star(t_{ij})$ the underlying process of interest. In functional data, the vector $\boldsymbol{y_i^\star} = \{y^\star(t_{ij})\}_{j=1,\dots,n_i}$ for unit i is considered as a random realization of an unknown underlying function. Consequently, the whole set of longitudinal trajectories $\boldsymbol{y^\star} = (\boldsymbol{y_1^\star}, \dots, \boldsymbol{y_N^\star})^\top$ is a collection of realizations of an unknown random function f assumed to be smooth. This unknown function has mean $\mu(t)$ and covariance function $G(s, t) = \text{cov}(f(t), f(s))$. FPCA is essentially based on the Karhunen–Loève decomposition [27, 28] which states that the underlying unknown smooth function f can be decomposed into:

$$f_i(t) = \mu(t) + \sum_{k=1}^\infty \xi_{ik} \phi_k(t) \tag{2}$$

where the index i in f_i emphasizes the fact that the random realization of f is subject specific, ξ_{ik} and ϕ_k are the eigenvalues (or functional principal component scores) and eigenfunctions from the decomposition of the covariance operator G, respectively. From this decomposition, dimension reduction is straightforward by limiting the sum in (2) to the first K components only:

$$f_i(t) = \mu(t) + \sum_{k=1}^{K} \xi_{ik} \phi_k(t) \qquad (3)$$

From the noisy data y_{ij}, local linear smoothers [10, 12] are used to estimate the mean function μ and the covariance function G. The estimated covariance surface is discretized to make spectral decomposition of the covariance operator easier and from that, eigenfunctions are obtained [11, 29]. In the case of dense functional data, the scores can be estimated using a numerical integration [10] which is not feasible in the case of sparse data. Functional principal components analysis through conditional expectation (PACE) [12] has been proposed as an alternative way of computing the eigenvalues in case of sparse data. Using properties of conditional Gaussian multivariate density, an expression for $\mathbb{E}(\xi_i | Y_i)$ is obtained, very close in spirit to what is done in linear mixed-effect models. From that, by plugging in the estimated eigenvalues, eigenfunctions, covariance, and mean function, an estimate of the score can be computed. In PACE, the scores are assumed to be Gaussian but the method is robust to violation of this assumption [12].

When using FPCA, the determination of K in (3) is critical. Apart from rare exceptions, K is determined from the data by optimizing a statistical criterion such as percentage of variance explained, AIC, or BIC where cross-validation can be used. Fitted values of existing observations or predictions of new data can be easily computed by plugging into (3) the estimated $\hat{\mu}(t_{ij})$ and $\hat{\phi}_k(t_{ij})$, and the individual eigenvalues $\hat{\xi}_{ik}$ computed using the PACE algorithm.

Many R packages have implemented FPCA for functional data but only a few are suited to the case of sparse and irregular functional data, in particular during the prediction step for new units. For example, fda [30] does not implement PACE and requires the same observation grid for all units which is not realistic in many health contexts. Package MFPCA [31] is suited to sparse and irregular functional data but it struggles to make predictions on new units with different observation grids. We identified two packages for the estimation of FPCA in case of sparse and irregular data: face [32] and fdapace [33]. Their major difference lies in the specification of the smoothing used to estimate the functional components: fdapace uses local weighted polynomial smoothing while face uses P-splines smoothing. We thus considered the two implementations in our work.

Contrarily to linear mixed-effect models, no theoretical justification exists regarding FPCA robustness to missing at random data, and no empirical study has explored this issue so far.

4 | Simulation Study

We numerically evaluated the performances of FPCA to analyze sparse, noisy and irregular repeated data in the presence of dropout with (i) a comparison of the predicted values with those obtained by mixed models considered as the gold standard approach, and (ii) an evaluation of the robustness of FCPA estimations to dropout. We followed the recommendations of Morris, White, and Crowther [34] to plan and report these two simulation studies.

4.1 | Comparison Between FPCA and MM

4.1.1 | Aims

The first simulation aimed to contrast the performances of MM and FPCA to predict missing values under different scenarios of dropout with different implementations of MM and FPCA.

4.1.2 | Data Generation Mechanism

We generated individual repeated data in samples of $N = 700$ participants. The sparse visit process was generated from a fixed grid of visits between $t = 0$ and $t = 12$ with three levels of sparsity in the visits times: a measure every 3, 2, or 1 years leading to a maximum of 5, 7, or 13 observations per individual.

To mimic a more realistic observation window, the individual visit times t_{ij} were obtained by adding a random uniform noise around the theoretical visits for each observation (± 1.5, 1, 0.5 years for visits every 3, 2, and 1 years, respectively). This uniform noise adds a first level of irregularity to the observation grid by making it specific to each individual. The repeated marker data at each visit time t_{ij} were generated according to a mixed model with a nonlinear shape over time captured by a five-parameter logistic model as implemented in the nlraa package [35]:

$$y_i^\star(t_{ij}) = y_{\infty i} + \frac{y_{0i} - y_{\infty i}}{(1 + \exp(\alpha_i \times (\log t_{ij} - \log \tau_i)))^{\theta_i}}$$

where $y_{0i} = 10 + b_{i1}$ and $y_{\infty i} = 35 + b_{i2}$ are the individual-specific asymptotic values of y for 0 and infinite times, respectively; $\tau_i = 10 + b_{i3}$ is the time when the mean value of y is attained (only if $\theta_i = 1$), that is, $y(\tau_i) = (y_{0i} + y_{\infty i})/2$, $\alpha_i = 5 + b_{i4}$ controls the steepness of the transition and $\theta_i = 1 + b_{i4}/3$ is an asymmetry parameter. Individual-specific trajectories were generated assuming random-effects $b_i = (b_{1i}, b_{2i}, b_{3i}, b_{4i})^\top \sim \mathcal{N}(0, B)$ with B a diagonal 4×4-matrix with diagonal $(3, 5, 2, 2)^\top$. The prone-to-error observations were derived by adding an independent homoscedastic Gaussian error $\varepsilon_{ij} \sim \mathcal{N}(0, 3^2)$ such that: $y(t_{ij}) = y_i^\star(t_{ij}) + \varepsilon_{ij}$.

At this stage, we have a complete set of repeated measures for each participant up to the administrative censoring at year 12. We added a second level of irregularity to the observation grid by truncating the observation process at an individual time of dropout (i.e., early leave from the study), making the observation grid length subject-specific. We considered a dropout with two intensities (30% and 60%) and the six following missing data scenarios (see supplementary material for details):

- MCAR: the participant drops out at t_{ij} with a probability determined by a logistic model with t_{ij} as predictor;
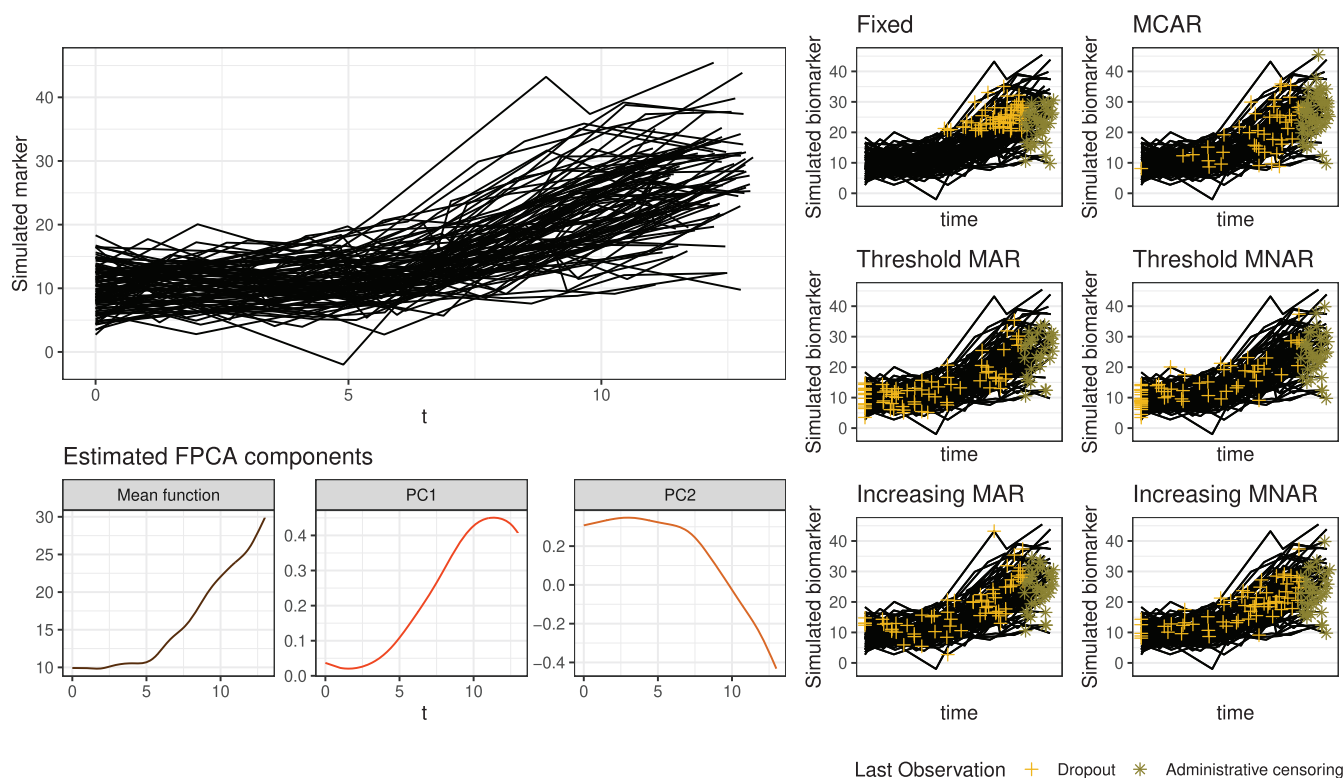
FIGURE 2 | Top-left and right panels: Generated individual longitudinal data according to a mixed model for $n = 100$ randomly selected participants prior to dropout (top-left) and after six different generated mechanisms of dropout with rate 0.6 (right). Bottom-left: The mean function and the two functional principal components that can be derived from the top-left individual trajectories using a functional principal component analysis (FPCA).

- fixed MAR: the participant systematically drops out at t_{ij} if $Y_{i(j-1)}$ is above a threshold ν;

- threshold MAR: the participant drops out at t_{ij} with a probability determined by a logistic model with the indicator that $Y_{ij-1} > \nu$ as predictor;

- increasing MAR: the participant drops out at t_{ij} with a probability determined by a logistic model with Y_{ij-1} as predictor;

- threshold MNAR: the participant drops out at t_{ij} with a probability determined by a logistic model with the indicator that $Y_{ij} > \nu$ as predictor;

- increasing MNAR: the participant drops out at t_{ij} with a probability determined by a logistic model with Y_{ij} as predictor;

The data generated as described above satisfies the three main challenging characteristics of longitudinal data usually encountered in epidemiological studies: sparsity with three levels of visit frequency, error-proneness with a Gaussian noise on the repeated measures of the marker, and irregular time grid both in visit occurrence and grid length. The generated longitudinal data of 100 randomly selected individuals are displayed in Figure 2. The right panels illustrate the two types of early truncation of the longitudinal process: the dropout under the above mechanisms (with a rate of 0.6) and the administrative censoring at year 12.

4.1.3 | Estimands

Each sample was separated into a training set of 200 participants and a test set of 500 participants. The estimands of interest were the predictions of the marker values y_{ij} in the test set whether they were observed or missing (true generated values were known from the generation procedure). These predictions were computed from the estimations of FPCA and MM on the training set.

4.1.4 | Methods

For each simulation setting, that is for a fixed frequency of visits (every 3, 2, and 1 years), for a fixed rate of dropout (30% or 60%), for a fixed missing data scenario (MCAR, fixed MAR, threshold MAR, increasing MAR, threshold MNAR, increasing MNAR), we estimated the following models:

- Linear mixed-effect models (including a random effect on the intercept and on each time function) with time trend: modeled as a quadratic trend (LMM_quad); modelled as a cubic trend (LMM_cub); approximated by natural splines with two internal knots at the terciles (LMM_spl_quant); approximated by natural splines with two equidistant internal knots (LMM_spl_equi).

- Functional principal component analysis with the number of principal components chosen such that the percentage of explained variance reached 90% (FPCA_fve90_fdapace) or 99% (FPCA_fve99_fdapace) in `fdapace` or 99% (FPCA_face_fve99) in `face`.

- Shared random effect joint model (JM) in which the longitudinal submodel followed a quadratic time trend (random intercept and random slope on each time function) and the survival submodel was a proportional hazard model with

a baseline risk function approximated by B-splines (three internal knots at the quartiles) and a dependence through the current underlying value of the marker. Note that the joint model was only estimated in the two MNAR scenarios (threshold MNAR and increasing MNAR) as a *gold standard* comparator for the MNAR situation.

The linear mixed-effect models were estimated using the `lcmm` package [36] (function `hlme`) and predictions were obtained using the `predictRE` function. The joint models were estimated using the `JM` package [37], and predictions computed with the `predict` function. FPCA models were estimated with the `fdapace` package [33] and the `face` package [32]. With fdapace, prediction was only possible for times available in the training set. We thus discretized the time window (with a step of 0.5) to ensure that every discretized t_{ij} in the test set also appeared in the training set, and used the internal `predict` function of `fdapace` and `face`.

4.1.5 | Performance Measure

For each setting, the procedure was replicated on 1000 replicates. The performances were evaluated with the root mean square error (RMSE) by separating: (i) the RMSE computed on the actual observations of the test set (i.e., non-missing due to dropout), and (ii) the RMSE computed on the missing data of the test set (after time of dropout).

To make all the RMSEs comparable across models within a fixed scenario, we considered standardized RMSEs. To do so, we first estimated a reference flexible model in the complete training set (i.e., without any missing data). This reference model was chosen as a mixed model with B-splines time trend with three internal knots placed at the quartiles. We computed the RMSE of this model on test data (either missing data only, or observed data only) and defined it as the reference RMSE. Then, for every other model estimated on the training set with missing data, we computed their RMSEs on test data (either missing data only, or observed data only) and standardized them by dividing them by the corresponding reference RMSE.

4.1.6 | Results

The standardized RMSEs of each model computed on the missing test data are reported in Figure 3 (see Figures S2, S3, and S4 for the observed test data) for four missing data scenarios (MCAR, fixed MAR, increasing MAR and increasing MNAR), two rates of dropout (30% and 60%), and the three frequencies of sparse visits.

An overall look at mixed models RMSEs shows that flexible mixed models, those based on spline parametrization, rank systematically among the best for every missing data scenario, dropout rate and observation grid. As expected, virtually the same standardized RMSE were obtained for MM and FPCA in the MCAR case, whatever the frequency of visit and the rate of dropout. With a fixed MAR dropout after an observation above a given threshold, the MM relying on splines showed the lowest standardized RMSEs. With a 30% rate of dropout, those of FPCA were comparable or even lower when data were denser (every year). However, with a 60% rate of dropout, the

performances of FPCA dropped, probably due to the sudden and local loss of information. With a MAR dropout that increases as the last observed marker value increased, the dropout is more spread along the observation window and FPCA lead to very competing standardized RMSEs compared to MM with splines time trends. This is true whatever the rate of dropout, and even provide slightly better fit for denser data and lower rate of dropout, underlying their flexibility to capture complex trends. With a MNAR dropout that increases as the simultaneous marker level increases, MM and FPCA models show larger standardized RMSEs compared to the more appropriate JM. However, again FPCA shows worse behavior than MM when the rate of dropout is high (60%). The two implementations of the FPCA gave overall the same results. We observed an expected but very tiny improvement when targeting 99% of explained variance rather than only 90%. The different specifications of the components between `face` and `fdapace` gave overall the same results in all scenarios. The fit of the spline-based FPCA (in `face`) was slightly better than the one of the local-polynomial based FPCA (in `fdapace`) when the dropout was systematic after the outcome reached a certain level (fixed MAR) but the differences remained negligible in all the other scenarios.

In summary, this simulation study suggests that, with sparsely and irregularly measured data, FPCA behave similarly as the flexible linear mixed-effect models in MCAR, MAR, and MNAR scenarios. In some specific situations when combining a higher rate of dropout and very specific missing scheme (systematic dropout above a threshold), the FPCA seemed to perform slightly worse in capturing the entire underlying pattern. These results suggest that FPCA may constitute a suitable alternative to MM when the information loss is progressive, allowing the FPCA to retrieve the underlying trend.

The results for the threshold MAR and threshold MNAR cases, displayed in Figure S1, lead to the same conclusions.

4.2 | Robustness of FPCA to Dropout

4.2.1 | Aims

This second simulation study aimed to evaluate to which extent the components and scores of FPCA were correctly estimated in the presence of dropout.

4.2.2 | Data Generation Mechanism

The data generation mechanism was exactly the same as in the first simulation study except that the samples were of size $N = 200$ and the time function for the underlying marker trajectory was generated according to a two-component FPCA rather than a mixed model.

The generating mean function and principal components, displayed in the bottom-left panel of Figure 2, were chosen as those obtained after estimating a FPCA on a dataset simulated as in Section 4.1. For each individual, we then randomly sampled Gaussian scores $\xi_i \sim \mathcal{N}(0, \Xi)$, (with Ξ taken as the empirical variance of the scores derived from the FPCA) and applied Equation (3) to generate the functional data at each observed visit time.
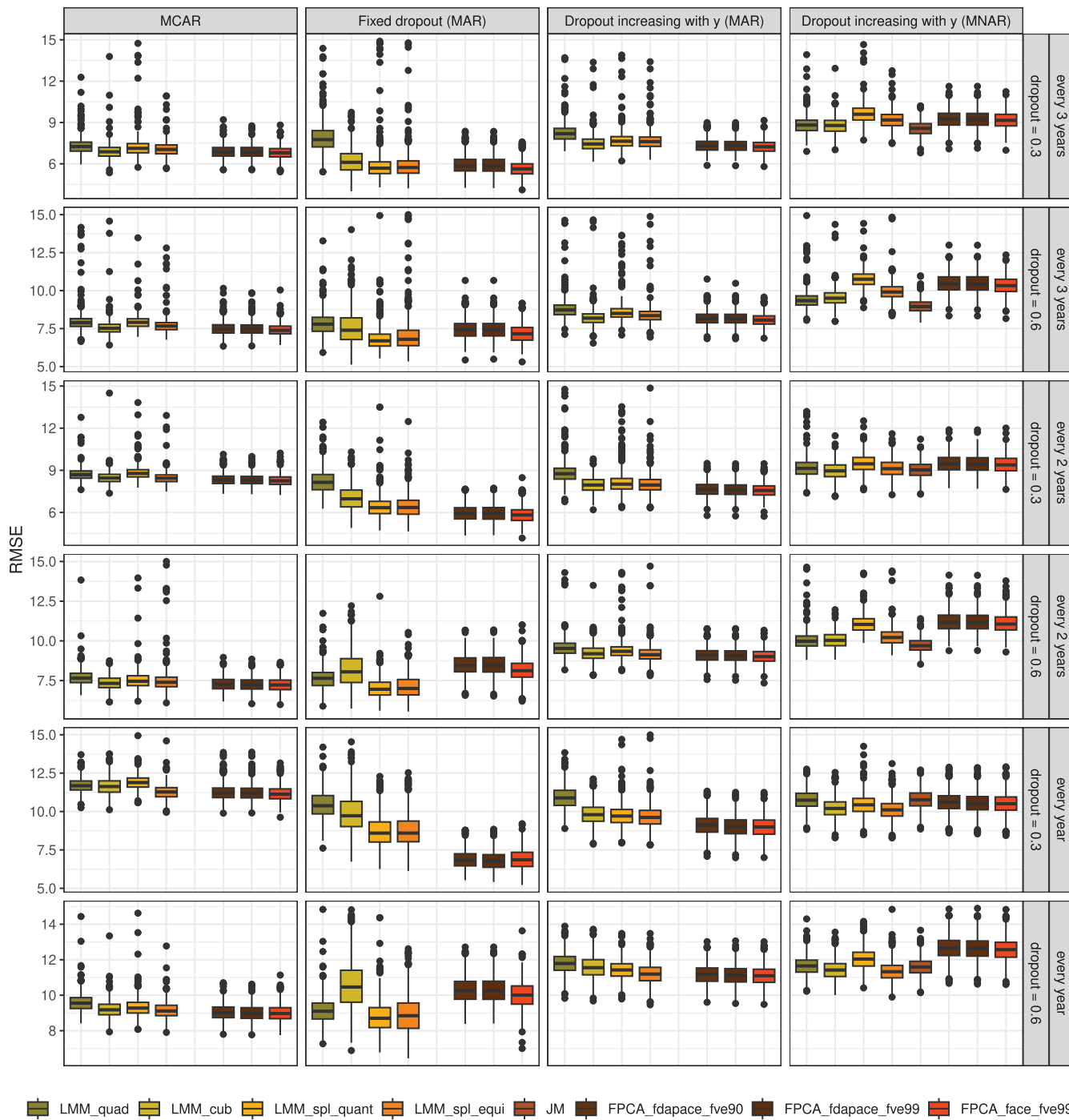
FIGURE 3 | Standardized RMSEs computed on missing test data over 1000 replicates for four missing data scenarios (MCAR, fixed MAR, increasing MAR and increasing MNAR from left to right), two rates of dropout (30% and 60%) and three frequencies of sparse visits (every 3, 2, and 1 year from top to bottom). Each boxplot summarizes the RMSEs of a model relative to the one of a reference model estimated on complete data (from left to right): LMM_quad, LMM_cub, LMM_spl_quant, LMM_spl_equi, JM (only for the MNAR case), FPCA_fdapace_fve90, FPCA_fdapace_fve99, FPCA_face_fve99.

4.2.3 | Estimands

The quantities of interest were the estimated mean function and the estimated functional principal components.

4.2.4 | Methods

For each setting (rate of dropout, mechanism of dropout and frequency of visits), we applied a FPCA implemented in `fdapace`

package [33] to the observed data, fixing the number of components to 2.

4.2.5 | Performance Measure

We evaluated the relative bias between the estimated mean function or principal components, and their true counterparts on a fixed time grid over 1000 replicates. Additionally, we visually

compared the empirical distribution of the estimated mean function and functional principal components to the generated ones.

4.2.6 | Results

The simulation results are reported in Figure 4 for both the MCAR and increasing MAR scenarios when considering a measure every 2 years. As expected, in the MCAR scenarios (upper panels), there is no bias in the estimations whatever the dropout rate (0%, 30%, or 60%) and the estimated curves are close to the true generated mean and principal components. In the increasing MAR (lower panels) and the threshold MAR (Figure S12) scenarios, the principal components are overall estimated without bias whatever the dropout rate (0%, 30%, or 60%) and the variability of the estimates does not substantially change with the dropout rate. The mean function in contrast shows a small bias at the end of the observation window that increases as the rate of dropout increases. Indeed the mechanism of dropout imposes a larger proportion of missing data by the end of the observation window. These observations are further illustrated by the 10 randomly selected estimated curves which highlight the departure from the true mean function in the 30% and 60% dropout cases. With a fixed MAR dropout after an observation above a threshold, the FPCA performed poorly at the end of the observation window as expected due to the sudden and systematic loss of information. In the increasing and threshold MNAR scenarios (Figures S13 and S14), the relative bias for the mean function is overall larger and especially at the end of the follow-up. The same results (Figures S5–S10 and S15–S20) were observed for the other frequencies of observations.

5 | Application to Cognitive Trajectories in Prodromal Dementia

We applied FPCA and compare it to MM to flexibly describe the trajectories of the four cognitive functions (global functioning, visual memory, verbal fluency, executive functioning) before clinical dementia, and contrast them with those of matched controls in the case-control sample nested within the 3C study. Each cognitive function was measured by a specific score: MMSE for global functioning, BVRT for visual memory, IST for verbal fluency and TMTA for executive functioning. The main characteristics of the nested case-control study sample are reported in Table 1 and individual trajectories and missingness mechanisms are plotted in Figure 1.

The analytical sample comprised 74% of women in both groups; 61% had at least finished secondary school and 25.3% were carriers of at least one APOE-ε4 allele. The median age at diagnosis was 84.8 (interquartile range [IQR] = 81.6 − 88.7) for cases. The matched index age for controls was 84.8 (IQR = 81.6 − 88.8). The four cognitive scores distribution implies a worst cognitive condition for cases with a median systematically lower and an interquartile range extending systematically below those of controls. The number of repeated measures per participant was similar between cases and controls and also among cognitive scores except for TMTA, which had fewer measures.

A separate analysis was conducted among cases and controls. For each subset and each cognitive score we estimated: a FPCA

with a targeted percentage of explained variance of 99%, and a linear mixed-effect model in which the cognitive trajectory was modeled with natural cubic splines with two internal knots at the quartiles, and boundary knots at −10 and 2 years to reduce the influence of observations beyond this range. With FPCA as implemented in fdapace, no control was possible for reducing the influence of time periods during which the information is sparser. For both methodologies, 95% confidence intervals were computed: directly from the lcmm output based on the inverse of the negative Hessian matrix for the mixed models, and computing bootstrap confidence intervals with 500 bootstrap samples with the function GetMeanCI for the FPCA.

The estimated marginal trajectories of the cognitive markers obtained by the two methods in each case/control group are displayed in Figure 5. Overall, the estimated mean trajectories and 95% confidence intervals were very similar across methods except for slight differences observed at the end of the observation window. The predicted mean trajectories obtained from the linear mixed-effect model were however much smoother than those estimated by the FPCA.

The mean trajectories of the four cognitive functions (visual memory, verbal fluency, global functioning, and executive functioning) showed very different trends for controls and cases. Controls had a slow linear decline over time for all the cognitive functions. Far from dementia diagnosis, cases also had a slight linear decline, almost overlapped with the one of the controls for global functioning, very close for visual memory, and almost parallel but approximately 0.1 points and 5 points below for executive functioning and verbal fluency respectively. Then closer to diagnosis, each cognitive decline showed a sharp acceleration among cases: around 5 years before diagnosis for verbal fluency, 3 years before diagnosis for executive functioning and 2.5 years before diagnosis for visual memory and global functioning.

6 | Discussion

Initially developed for dense and regular functional data, FPCA has been recently extended to handle sparsity [12]. If FPCA can be used to impute missing data, the robustness of FPCA to the presence of missing data remained uncertain. In this manuscript, we have conducted an empirical simulation study to evaluate the robustness of FPCA in the presence of missing data caused by dropout. Our findings demonstrate that, in practical applications, FPCA is well-suited for analyzing longitudinal data, even when missing data follow a missing at random (MAR) mechanism, exhibiting comparable performance to linear mixed-effect models known for their robustness in such scenarios. This observation holds true across various MAR scenarios, observation grid densities, and dropout rates with exception for specific scenarios with the combination of very frequent and systematic dropout above a threshold. FPCA along with linear mixed-effect models do not seem to be robust to missing not at random data. These results are based on two FPCA implementations: fdapace and face, but from our experience,
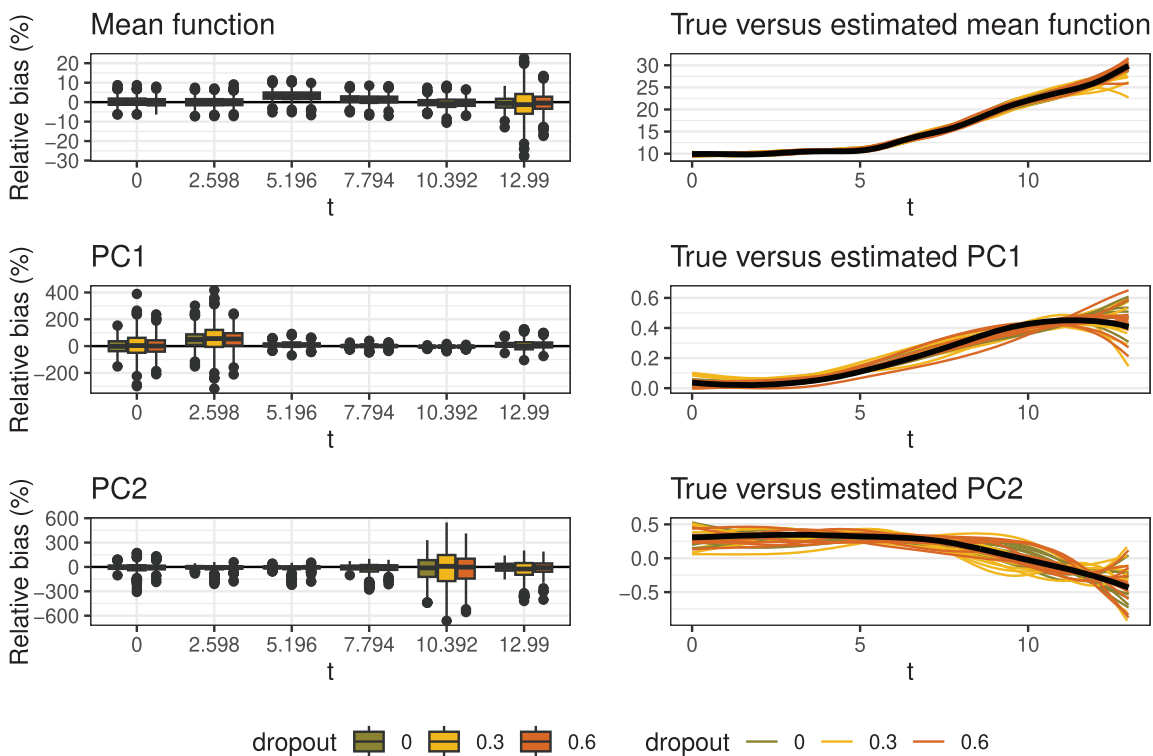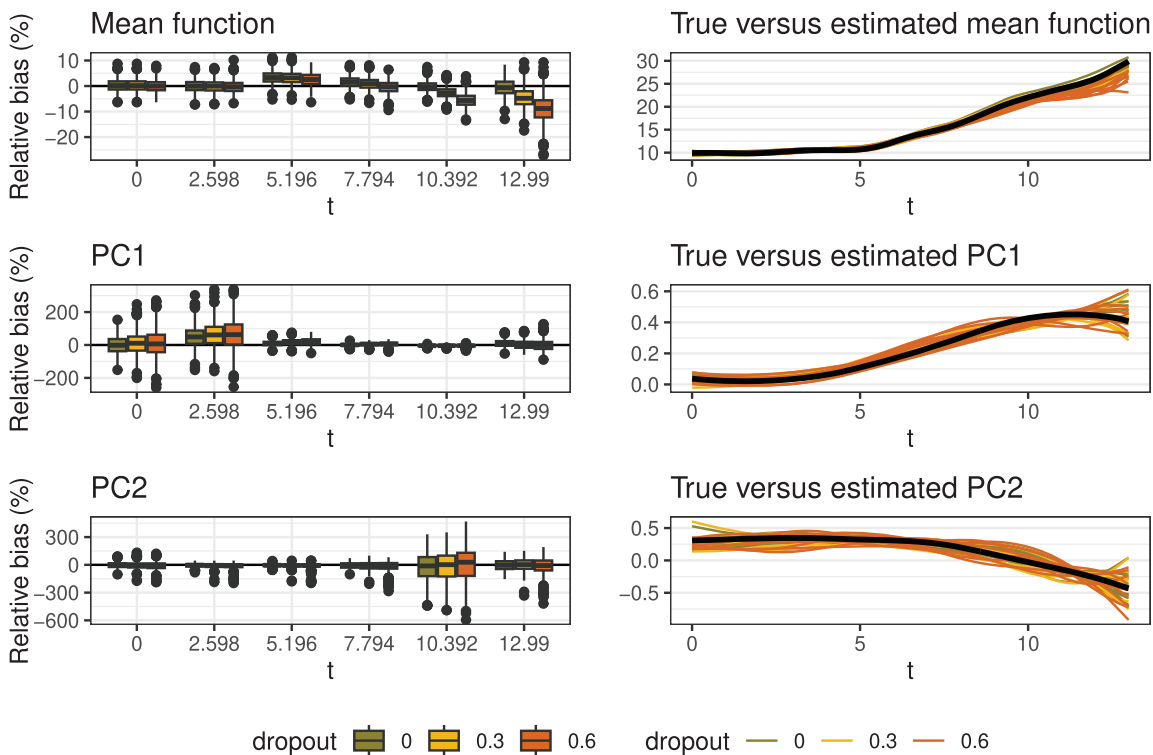
FIGURE 4 | Relative bias over 1000 replicates (left panel), and true plus 10 randomly estimations (right panel) for the mean function and each of the two functional principal components in MCAR and MAR scenarios. The settings are a measure every 2 years for 12 years with a dropout rate of 0 (no dropout), 30% or 60%.

TABLE 1 | Summary statistics of demographic variables and cognitive scores in the case and control groups from the nested matched case-control ($N = 348$, the 3C study cohort). Are reported count (frequency) for categorical variables and median (interquartile range) for continuous data and number of measures.

Variable	Cases ($N = 174$)	Controls ($N = 174$)
Women	129 (74.1%)	129 (74.1%)
Index age*	84.8 (81.6 − 88.9)	84.8 (81.6 − 88.7)
Secondary school or higher	106 (60.9%)	106 (60.9%)
At least one APOE-ε4 allele	44 (25.3%)	44 (25.3%)
Follow-up at diagnosis (in years)	9.72 (6.78 − 12.2)	9.70 (7.08 − 12.4)
MMS score at index age	23 (21 − 25)	28 (25 − 29)
Number of individual measures of MMS	6 (4 − 7)	6 (5 − 7)
BVRT score at index age	9 (8 − 11)	11 (9 − 12)
Number of individual measures of BVRT	5 (4 − 6)	5 (4 − 7)
IST30 score at index age	28 (22 − 34)	39 (34 − 47)
Number of individual measures of IST30	5 (4 − 7)	6 (4.25 − 7)
TMTA score at index age	0.25 (0.21 − 0.35)	0.39 (0.31 − 0.48)
Number of individual measures of TMTA	4 (3 − 5)	4 (3 − 5)

*Index age corresponds to the matching age at which cases were diagnosed with dementia, and controls were still free of dementia.
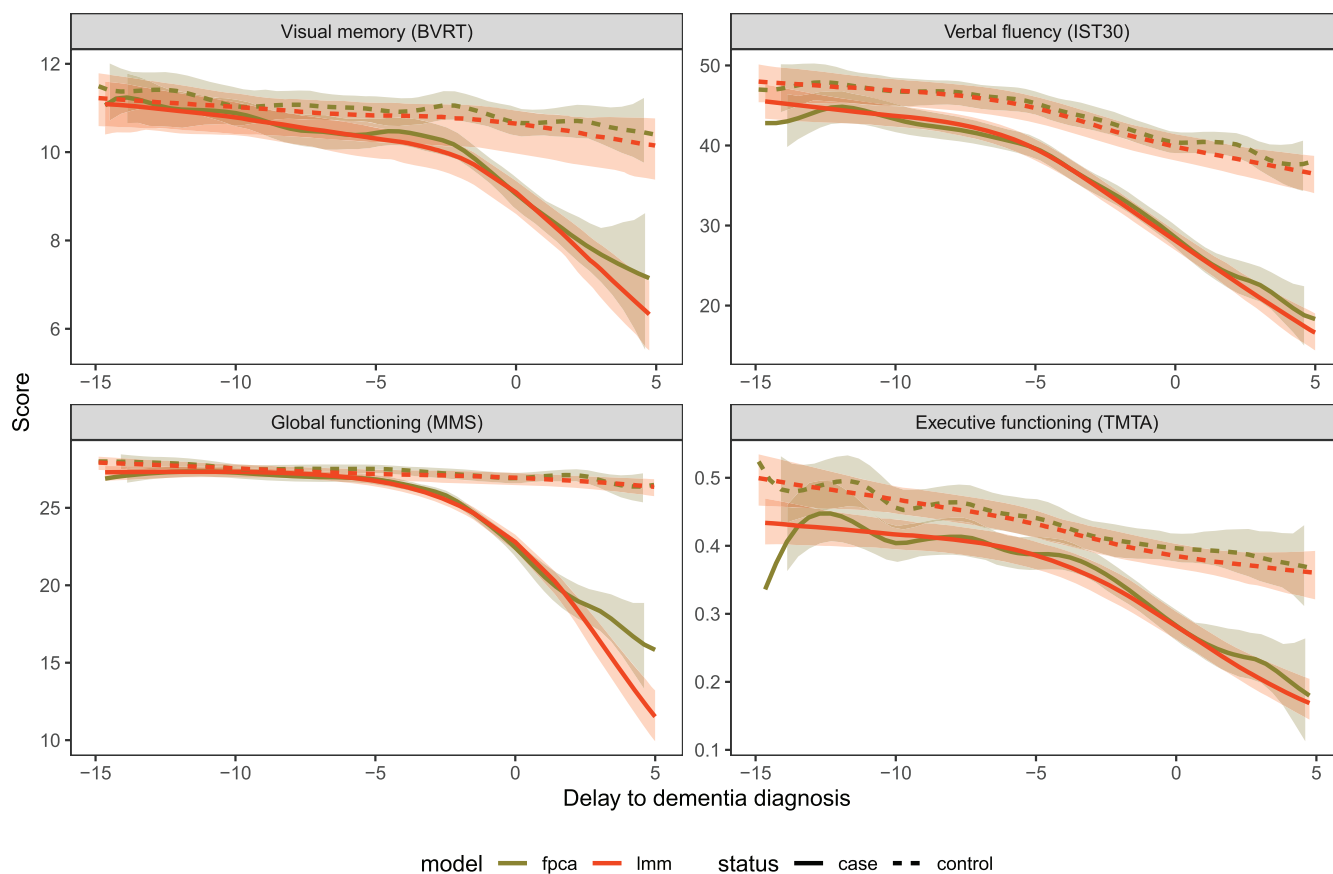

FIGURE 5 | Marginal estimation from the linear mixed-effect model and mean function estimation from the functional principal component analysis (FPCA) and their 95% confidence intervals for each cognitive marker among cases and matched controls of the nested case-control study from 3C cohort ($N = 348$).

fdapace is currently the easiest implementation to use with sparse and irregular functional data.

This empirical evaluation specifically targeted dropout, a particular category of missing data pattern wherein the absence of data at any point implies that all subsequent visits are also missing. The dropout scenario may induce a larger information loss than intermittent missing data where information from the longitudinal trajectory can be retrieved, even at the end of the window time. Hence, we anticipate the robustness to dropout found in this work to also apply to intermittent missing data. Note that it is important to distinguish missing data due to dropout and missing data truncated by death [38, 39]. In this work, we only focused on missing data due to dropout.

Our work was restricted to designs in which the observation times were non-informative. This is the case in the 3C study as in most cohort studies in which the visits are planned beforehand. When observation times depend upon the measured outcome (e.g., when participants consult each time their health condition worsens), specific functional data techniques have been very recently proposed [40, 41], including the use of weights in the estimating equations of FPCA [40]. We used the standard weighting scheme implemented in PACE algorithm in which each observation has equal weight [12, 42]. Alternatively, subject-specific weights could be applied for balancing the contribution of each subject [40] in case of informative visit times. In the presence of informative dropout, no weighting technique has been developed so far in software to our knowledge. We leave for future works the study of weighted versions, including notably weighting schemes proposed in marginal models estimated by generalized estimating equations in presence of dropout or death [39].

In our application study, without any assumption on the shape of the cognitive trajectories, both methods (based on MM and FPCA) showed very similar trends. Despite the absence of control on the FPCA degree of smoothing using fdapace [33], the mean predicted trajectories were smooth over time except for the TMTA for which variations at the very beginning of the observation window suggested an overfitting (see Figure 5). Compared to other scores, TMT was not collected at the second visit of the cohort and fewer measures were available between −15 and −10 (see last row in Figure 1).

Cognitive trajectories exhibited a linear mean trend for controls throughout the window of interest. In contrast, the cognitive trajectories of cases, closely resembling those of controls distant from the diagnosis, exhibited a break with an accelerated decline as the diagnosis approached. This two-stage trend in pathological cognitive decline leading towards dementia had already been identified in cognitive ageing research [43, 44]. This empirical study highlights the usefulness of FPCA to closely describe, summarize and predict longitudinal data. As a non-parametric approach, it avoids dependence upon strong assumptions and can adapt to any shape of trajectory. However, inference tools are not straightforwardly available with FPCA contrarily to MM which benefits from the maximum likelihood framework. For instance, parametric change point mixed models have been proposed to conduct inferential tasks to specifically explore this cognitive decline two-stage trend [45, 46].

A parallel between the mixed-effect model and the FPCA frameworks exists. The role of the individual random effects in the mixed-effect model is very similar to the role of the individual scores in the FPCA. They both represent individual deviations to the mean trajectory: the marginal model in mixed-effect model and the mean function in the FPCA. Moreover, under Gaussian assumption, the random effect prediction for linear mixed-effect model is based on Gaussian conditional expectation properties and its formula is very close to FPCA score's formula from the PACE algorithm. Despite this parallel, FPCA and more broadly any method from functional data analysis were initially developed for dense and regular functional data. This work joins others that explored how to handle missing data using multiple imputation techniques in functional regressions with functional data as the covariate [47] or as the outcome [48]. All these techniques open up possibilities regarding the utilization of functional data analytical tools in the context of sparsely and irregularly measured data increasingly available in health studies.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

1. W. J. Boscardin, J. M. Taylor, and N. Law, "Longitudinal Models for AIDS Marker Data," *Statistical Methods in Medical Research* 7, no. 1 (1998): 13–27, https://doi.org/10.1177/096228029800700103.

2. D. Ilic, M. Djulbegovic, J. H. Jung, et al., "Prostate Cancer Screening With Prostate-Specific Antigen (PSA) Test: A Systematic Review and Meta-Analysis," *BMJ* 362 (2018): k3519, https://doi.org/10.1136/bmj.k3519.

3. A. Alpérovitch, P. Amouyel, J. F. Dartigues, et al., "Les études épidémiologiques Sur le Vieillissement en France: De l'étude Paquid à l'étude des Trois Cités," *Comptes Rendus Biologies* 325, no. 6 (2002): 665–672, https://doi.org/10.1016/S1631-0691(02)01476-2.

4. A. Lokku, L. S. Lim, C. S. Birken, E. M. Pullenayegum, and TARGet Kids! Collaboration, "Summarizing the Extent of Visit Irregularity in Longitudinal Data," *BMC Medical Research Methodology* 20, no. 1 (2020): 135, https://doi.org/10.1186/s12874-020-01023-w.

5. J. W. Hogan, J. Roy, and C. Korkontzelou, "Handling Drop-Out in Longitudinal Studies," *Statistics in Medicine* 23, no. 9 (2004): 1455–1497, https://doi.org/10.1002/sim.1728.

6. G. Verbeke, "Linear Mixed Models for Longitudinal Data," in *Linear Mixed Models in Practice: A SAS-Oriented Approach Lecture Notes in Statistics*, eds. G. Verbeke and G. Molenberghs (New York, NY: Springer, 1997), 63–153.

7. D. B. Rubin, "Inference and Missing Data," *Biometrika* 63, no. 3 (1976): 581–592, https://doi.org/10.2307/2335739.

8. R. J. A. Little and D. B. Rubin, *Statistical Analysis With Missing Data* (Hoboken, NJ: Wiley, 1987).

9. M. Davidian and D. M. Giltinan, "Nonlinear Models for Repeated Measurement Data: An Overview and Update," *Journal of Agricultural, Biological, and Environmental Statistics* 8, no. 4 (2003): 387–419, https://doi.org/10.1198/1085711032697.

10. J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*. Springer Series in Statistics (New York, NY: Springer, 2005).

11. J. L. Wang, J. M. Chiou, and H. G. Mueller, "Review of Functional Data Analysis," (2015), arXiv:1507.05135 [stat].

12. F. Yao, H. G. Müller, and J. L. Wang, "Functional Data Analysis for Sparse Longitudinal Data," *Journal of the American Statistical Association* 100, no. 470 (2005): 577–590.

13. M. F. Folstein, S. E. Folstein, and P. R. McHugh, ""Mini-mental state". A Practical Method for Grading the Cognitive State of Patients for the Clinician," *Journal of Psychiatric Research* 12, no. 3 (1975): 189–198, https://doi.org/10.1016/0022-3956(75)90026-6.

14. A. L. Benton, *The Revised Visual Retention Test: Clinical and Experimental Applications*, 3rd ed. (New York, NY: Psychological Corporation, 1963).

15. B. Isaacs and A. T. Kennie, "The Set Test as an Aid to the Detection of Dementia in Old People," *British Journal of Psychiatry: The Journal of Mental Science* 123, no. 575 (1973): 467–470, https://doi.org/10.1192/bjp.123.4.467.

16. R. M. Reitan, "Validity of the Trail Making Test as an Indicator of Organic Brain Damage," *Perceptual and Motor Skills* 8, no. 3 (1958): 271–276, https://doi.org/10.2466/pms.1958.8.3.271.

17. N. M. Laird and J. H. Ware, "Random-Effects Models for Longitudinal Data," *Biometrics* 38, no. 4 (1982): 963, https://doi.org/10.2307/2529876.

18. C. Proust-Lima, H. Amieva, J. F. Dartigues, and H. Jacqmin-Gadda, "Sensitivity of Four Psychometric Tests to Measure Cognitive Changes in Brain Aging-Population-Based Studies," *American Journal of Epidemiology* 165, no. 3 (2007): 344–350, https://doi.org/10.1093/aje/kwk017.

19. C. Proust-Lima, J. F. Dartigues, and H. Jacqmin-Gadda, "Misuse of the Linear Mixed Model When Evaluating Risk Factors of Cognitive Decline," *American Journal of Epidemiology* 174, no. 9 (2011): 1077–1088, https://doi.org/10.1093/aje/kwr243.

20. J. Pan and R. Thompson, "Gauss-Hermite Quadrature Approximation for Estimation in Generalised Linear Mixed Models," *Computational Statistics* 18, no. 1 (2003): 57–78, https://doi.org/10.1007/s001800300132.

21. G. Molenberghs and G. Verbeke, *Linear Mixed Models for Longitudinal Data*. Springer Series in Statistics (New York, NY: Springer, 2000).

22. P. Diggle and M. G. Kenward, "Informative Drop-Out in Longitudinal Data Analysis," *Journal of the Royal Statistical Society: Series C: Applied Statistics* 43, no. 1 (1994): 49–93, https://doi.org/10.2307/2986113.

23. D. Rizopoulos, *Joint Models for Longitudinal and Time-To-Event Data: With Applications in R* (Boca Raton, FL: CRC Press, 2012).

24. C. Thomadakis, L. Meligkotsidou, N. Pantazis, and G. Touloumi, "Longitudinal and Time-To-Drop-Out Joint Models Can Lead to Seriously Biased Estimates When the Drop-Out Mechanism Is at Random," *Biometrics* 75, no. 1 (2019): 58–68, https://doi.org/10.1111/biom.12986.

25. P. Besse and J. O. Ramsay, "Principal Components Analysis of Sampled Functions," *Psychometrika* 51, no. 2 (1986): 285–311, https://doi.org/10.1007/BF02293986.

26. H. L. Shang, "A Survey of Functional Principal Component Analysis," *AStA Advances in Statistical Analysis* 98, no. 2 (2014): 121–142, https://doi.org/10.1007/s10182-013-0213-1.

27. K. Karhunen, "Zur Spektraltheorie Stochastischer Prozesse," (1946).

28. M. Loève, "Fonctions Aléatoires à Décomposition Orthogonale Exponentielle," *La Revue Scientifique* 84 (1946): 159–162.

29. J. A. Rice and B. W. Silverman, "Estimating the Mean and Covariance Structure Nonparametrically When the Data are Curves," *Journal of the Royal Statistical Society: Series B: Methodological* 53, no. 1 (1991): 233–243.

30. J. Ramsay, "Fda: Functional Data Analysis," (2023), R package version 6.1.4.

31. C. Happ and S. Greven, "Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains," *Journal of the American Statistical Association* 113 (2018): 649–659, https://doi.org/10.1080/01621459.2016.1273115.

32. L. Xiao, C. Li, W. Checkley, and C. Crainiceanu, "Face: Fast Covariance Estimation for Sparse Functional Data," (2022), R package version 0.1-7.

33. Y. Zhou, S. Bhattacharjee, C. Carroll, et al., "Fdapace: Functional Data Analysis and Empirical Dynamics," (2022), R package version 0.5.9.

34. T. P. Morris, I. R. White, and M. J. Crowther, "Using Simulation Studies to Evaluate Statistical Methods," *Statistics in Medicine* 38, no. 11 (2019): 2074–2102, https://doi.org/10.1002/sim.8086.

35. F. Miguez, "Nlraa: Nonlinear Regression for Agricultural Applications," (2022), R package version 1.5.

36. C. Proust-Lima, V. Philipps, A. Diakite, and B. Liquet, "Lcmm: Extended Mixed Models Using Latent Classes and Latent Processes," (2023), R package version: 2.0.2.

37. D. Rizopoulos, "JM: An R Package for the Joint Modelling of Longitudinal and Time-To-Event Data," *Journal of Statistical Software* 35, no. 9 (2010): 1–33.

38. B. F. Kurland, L. L. Johnson, B. L. Egleston, and P. H. Diehr, "Longitudinal Data With Follow-Up Truncated by Death: Match the Analysis Method to Research Aims," *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* 24, no. 2 (2009): 211, https://doi.org/10.1214/09-STS293.

39. A. Rouanet, C. Helmer, J. F. Dartigues, and H. Jacqmin-Gadda, "Interpretation of Mixed Models and Marginal Models With Cohort Attrition Due to Death and Drop-Out," *Statistical Methods in Medical Research* 28, no. 2 (2019): 343–356, https://doi.org/10.1177/0962280217723675.

40. C. Weaver, L. Xiao, and W. Lu, "Functional Data Analysis for Longitudinal Data With Informative Observation Times," *Biometrics* 79, no. 2 (2023): 722–733, https://doi.org/10.1111/biom.13646.

41. G. Xu, J. Zhang, Y. Li, and Y. Guan, "Bias-Correction and Test for Mark-Point Dependence With Replicated Marked Point Processes," *Journal of the American Statistical Association* 119, no. 545 (2024): 217–231, https://doi.org/10.1080/01621459.2022.2106234.

42. X. Zhang and J. L. Wang, "From Sparse to Dense Functional Data and Beyond," *Annals of Statistics* 44, no. 5 (2016): 2281–2321.

43. H. Jacqmin-Gadda, D. Commenges, and J. F. Dartigues, "Random Changepoint Model for Joint Modeling of Cognitive Decline and Dementia," *Biometrics* 62, no. 1 (2006): 254–260, https://doi.org/10.1111/j.1541-0420.2005.00443.x.

44. H. Amieva, H. Mokri, M. Le Goff, et al., "Compensatory Mechanisms in Higher-Educated Subjects With Alzheimer's Disease: A Study of 20 Years of Cognitive Decline," *Brain: A Journal of Neurology* 137, no. Pt 4 (2014): 1167–1175, https://doi.org/10.1093/brain/awu035.

45. A. Dominicus, S. Ripatti, N. L. Pedersen, and J. Palmgren, "A Random Change Point Model for Assessing Variability in Repeated Measures of Cognitive Function," *Statistics in Medicine* 27, no. 27 (2008): 5786–5798, https://doi.org/10.1002/sim.3380.

46. C. Segalas, H. Amieva, and H. Jacqmin-Gadda, "A Hypothesis Testing Procedure for Random Changepoint Mixed Models," *Statistics in Medicine* 38, no. 20 (2019): 3791–3803, https://doi.org/10.1002/sim.8195.

47. A. R. Rao and M. Reimherr, "Modern Multiple Imputation With Functional Data," *Stat* 10, no. 1 (2021): e331, https://doi.org/10.1002/sta4.331.

48. A. Ciarleglio, E. Petkova, and O. Harel, "Elucidating Age and Sex-Dependent Association Between Frontal EEG Asymmetry and Depression: An Application of Multiple Imputation in Functional Regression," *Journal of the American Statistical Association* 117, no. 537 (2022): 12–26, https://doi.org/10.1080/01621459.2021.1942011.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.